Latent transition analysis and offending lifestyle specialisation in England and Wales

Brian Francis[1] Jiayi Liu[2] and Keith Soothill[3]

[1]Professor of Social Statistics
[2]Research Associate
[3]Emeritus Professor of Social Research

Department of Maths and Statistics, Lancaster University, UK

B.Francis@Lancaster.ac.uk

The Lancaster research programme

Substantive: to investigate changes in the patterns of criminal careers over the lifecourse, specifically involving the nature of offending

Methodological: To develop new methods for assessing changes in the nature of offending over time, and to take advantage of modern administrative datasets such as the England and Wales Offenders Index.

General approach has been to use latent class analysis.

This talk: a focus on specialisation and changes over time through latent transition analysis.

What is specialization?

At least two views:

Paternoster et al (1998) “ Specialization is the extent to which an offender tends to repeat the same specific offence or offence type on successive events”

Gottfredson and Hirshi(1990) “Versatility is where offenders commit a variety of criminal acts, with no strong inclination to pursue a certain criminal act or pattern of criminal acts to the exclusion of others”.

Specialization for G&H is thus is the opposite of versatility – where offenders have a strong inclination to exclude certain criminal acts.

The definitions are subtly different – Paternoster talks about staying within the same type of offence and refers to successive events, whereas Gottfredson and Hirshi offer a far broader definition.

Specialization in offending

A number of approaches have been proposed.

a) Forward specialization coefficients – construct a transition matrix between offence types at event *t-1* and event *t* (event could be court appearance, or arrest etc).. Measure the divergence from randomness in staying in the same offence category. Farrington et al(1986)
   FSC=0 random transition, no specialisation
   FSC=1 complete specialisation

b) Diversity indices (eg Piquero et al,1999; Sullivan et al ;2006). Measures the degree of versatility in the offence history of an individual over a fixed period of time. $D = 1 - \sum_i p_i^2$ where p_i is the proportion of convictions of type i.

c) Regression approach. Can prior offending of type X predict future offending of type X? If so, then there is evidence of specialization. (Deane, Armstrong and Felson, 2005)

FSC- Example from Tarling(1993)

Five categories - Violence and robbery, sexual, burglary, theft and fraud, other
Data collected on 2077 offenders sentenced at 18 crown court centres in 1986-1987.
Data pooled over first 12 convictions. Transition matrix of counts is

	Court appearance k+1					
Court appearance k	VR	S	B	TF	O	TOTAL
Violence+robbery	213	16	156	246	197	819
Sexual	8	28	21	43	13	112
Burglary	206	26	1109	877	361	2578
Theft, handling, fraud	346	69	934	1556	553	3458
Criminal damage, drugs,motoring, other	208	15	312	520	431	1585
TOTAL	981	154	2531	3241	1555	8452

So, for sexual crime we expect = 2.04 for the diagonal cell, but we observe 28. FSC= (28-2)/(112-2) = 26/110 = 0.24

Forward specialisation coefficients from Tarling Study

	FSC	s.e.
Violence+robbery	0.16	0.01
Sexual	0.24	0.01
Burglary	0.19	0.01
Theft, handling, fraud	0.11	0.01
Criminal damage, drugs,motoring, other	0.13	0.01

Sexual offending is more specialised than other types of offending but still shows a large amount of versatility. Standard errors can also be calculated and reported.

Criticisms of specialization approaches.

a) Forward specialisation calculations have no calendar time concept – adjacent court appearances can be separated by a couple of weeks or by years. Also principal offence problem -need to classify a court appearance or arrest for rape and violence as either a sexual or violent offence.

b) Diversity indices depend on number of categories chosen. They produce individual scores and score distributions but difference from randomness is often not examined.

c) Regression approach relies on choice of other variables also used to predict future offending of type .X.

Do these measure what we want to measure?
Pasternoster’s view is really driven by methodology and the use of the forward specialization coefficient.

Measurement of specialization

Both the traditional application of the forward specialization coefficient and the diversity index fail to engage with the Gottfredson and Hirshi definition.

> *"Versatility is where offenders commit a variety of criminal acts, with no strong inclination to pursue a certain criminal act or pattern of criminal acts to the exclusion of others"*

The regression approach also could fail to identify whether the absence of prior convictions of particular types is associated with future offending of another type.

Perhaps we need an alternative viewpoint.

An alternative concept – lifestyle specialization

Idea is that offenders will engage in certain activities from the menu of available offences but not others.

Their "menu choice" may in addition change over the lifecourse.

Moves away from the idea of the versatility of "cafeteria-style" delinquency - a term first proposed by Klein (1971) where offence choice is "random", to a recognition that some metaphorical diners are vegetarian, some only eat chicken etc.

Thus some burglars will avoid people – burgle commercial premises and empty houses – and will be unlikely to engage in violence but may also handle stolen property. Other burglars might well relish the chance of confrontation when burgling houses and will become involved in violence and sexual offending.

There is some justification for this from interview studies and biographies of offenders. Can we find evidence in data?

Our approach

a) to identify a set of criminal lifestyles over the criminal career for a large group of offenders – finding patterns of offending.

b) To examine changes in criminal lifestyles by looking at transitions between criminal lifestyles at fixed transition points.

c) To examine the diversity of the lifestyles, identifying which lifestyles show greater versatility and which exhibit greater specialization.

d) To explore the reasons for diversity changes over the lifecourse

The methodology

We use latent transition analysis, taking offending over three broad time intervals – early teenage, late teenage and early 20s.

Latent transition analysis will identify offending patterns or typologies that co-occur in the dataset, and also estimate how offenders transit between offending patterns and also transit into non-offending.

We assume that these typologies are static, not dynamic. In other words, we assume that bicycle stealing and shoplifting ~(if a real typology) will co-occur in all age groups but with differing frequencies – it will not morph into bicycle stealing and (say) criminal damage.

The Offenders Index data set

We use the England and Wales Offenders Index – a Home Office research data set, which is a court based record of the criminal histories of all offenders in England and Wales from 1963 to the current day.

We analyse data from the Offenders Index Cohort study, which makes available six birth cohorts born in 1953, 1958, 1963, 1968, 1973, 1978 and followed through to 1999.

The birth cohorts give an approximate 1 in 13 sample of all offenders and samples all offenders born in the same four selected weeks for each cohort.

The index stores dates of conviction, the offence code of the conviction (very detailed) and the disposal or sentence.

We simplify the data, reducing the ~2000 offence codes to 38 major offences, after combining categories (Francis et al, 2004 EuroJCrim).

The data

We initially look at three time points with the female conviction data. Two transitions – one at age 15 and the second at age 20.

	Age							No. of female offenders in cohort
Birth Cohort	10-15	16-20	21-25	26-30	31-35	36-40	41-45	
1953								2217
1958								2348
1963								2569
1968								1797
1973								1071
1978								665
No. of female offenders in age group	2555	4659	3,132					10667

The 38 broad offence groups

1	Lethal violence (including attempts)		20	Theft (in a dwelling)
2	Violence		21	Theft (machines/meters/electricity)
3	Firearms/dangerous weapon (possession etc)		22	Theft from vehicles
4	Resisting arrest etc		23	Theft of vehicles
5	Kidnapping/false imprisonment		24	Attempted theft of/from vehicle
6	Sexual 16+		25	Shoplifting
7	Sexual under 16		26	Fraud and forgery
8	Sexual consensual		27	Receiving and handling
9	Prostitution		28	Criminal damage
10	Burglary (dwelling)		29	Drugs (possession etc only)
11	Aggravated burglary (dwelling, other)		30	Drugs (supply, including possession with intent)
12	Burglary (other)		31	Drugs (import/export/production)
13	Going equipped		32	Absconding/bail/breach offences
14	Robbery		33	Public order
15	Blackmail		34	Perjury/attempting to pervert course of justice
16	Vehicle taking (aggravated etc)		35	Dangerous Driving
17	Theft		36	Immigration
18	Theft from person		37	Child cruelty etc
19	Theft by employee		38	Other

Data analysis

Use binary indicators on the 38 broad offence groups within three five year age windows (11-15, 16-20, 21-25)

Define set of indicator variables within an age group and offender,

$O_{ija} = 1$ if offender *i* is convicted for offence *j* in age group *a*

$O_{ija} = 0$ otherwise.

(10-15)	(16-20)	(21-25)	
001000010000000	000000100000000	000000000000000	case 1
000000100000100	010010000100001	011001010000101	case 2

Five birth cohorts analysed 1953 1958, 1963, 1968, 1973 and look at female offending.

Calculate diversity index for each case and age group where offending happens.

Calculate average diversity over cohorts and age groups.

Offence diversity across age and cohort for female offenders

Birth Cohort	Age		
	10-15	16-20	21-25
1953	0.068	0.130	0.110
1958	0.116	0.144	0.157
1963	0.159	0.186	0.185
1968	0.141	0.190	0.224
1973	0.206	0.244	0.286

In general, increasing diversity with increasing age within each cohort.
Plus increasing diversity for more recent cohorts.

Or perhaps a calendar year effect – diversity increases with calendar year.

We propose a model where offending for each age and cohort is a mix of different offender lifestyles – some diverse and others specialised.

Changing proportions of these lifestyles will account for the observed changes in diversity. Latent transition analysis will provide the methodology.

Latent Class Analysis

For fixed window size and position, we define $\mathbf{O}_i$ to be the prevalence vector for offender *i* over the offences.
Assume there are *K* classes, with $k=1\ldots K$.

Let $\pi(k)$ be the probability of membership of class *k*, and p_{jk} the probability that there is at least one offence of type *j* given that the offender belongs to class *k*.

Then the likelihood is

$$L = f(\mathbf{O}) = \prod_i \sum_k \pi(k) p(\mathbf{O}_i | k)$$

where

$$p(\mathbf{O}_i | k) = \prod_j p_{jk}^{O_{ij}} (1 - p_{jk})^{1-O_{ij}}$$

Conditional independence given class membership.

We wish to fit a model where the latent classes are estimated globally over all individuals and ages, but the data points represent local events in the neighbourhood of age *a*.

The posterior probability of class membership will vary by age.

We extend the definition of the prevalence matrix to be O_{ija}

$O_{ija} = 1$ if offender *i* is convicted for offence *j* within the offence strip *a*– the window of width *h* years centred on age *a*

$O_{ija} = 0$ otherwise.

e.g.

Offence j	1	2	3	4	5	6	7	8	9	..	..	..	37	38
Age 16-20	0	0	1	1	0	0	0	1	0				0	0

With k classes, the latent class model then becomes:

$$L = f(\mathbf{O}) = \prod_i \prod_a \sum_k \pi(k) p(\mathbf{O}_{ia} | k)$$

where

$$p(\mathbf{O}_{ia} | k) = \prod_j p_{jk}^{O_{ija}} (1 - p_{jk})^{1 - O_{ija}}$$

Posterior probabilities

We can obtain the probabilities of an age strip belonging to a latent class.

$$q_{ika} = \frac{\pi(k)\prod_{j}\left(p_{jk}\right)^{O_{ija}}\left(1-p_{jk}\right)^{1-O_{ija}}}{\sum_{k=1}^{K}\pi(k)\prod_{j}\left(p_{jk}\right)^{O_{ija}}\left(1-p_{jk}\right)^{1-O_{ija}}}$$

An empirical transition matrix can then be obtained by summing over cases the product of the probability of being in one latent class *c* at time *a-1* and another latent class *d* at time *a*, and dividing by the sample size n. A typical cell estimate for the probability of moving from latent class *c* to latent class *d* at time *a* (t^{a}_{dc}) would be:

$$t_{dc}^{a} = \frac{\sum_{i} q_{ic(a-1)} q_{ida}}{n}$$

But can we estimate these as part of the model?

Modelling transition probabilities – Latent markov models

(joint work with Bartolucci and Pennoni – JRSSA , 2007)

Markov models provide a way forward to model both the transition probabilities and the latent classes. However, the number of parameters increases further.

Univariate markov model – based on work by Bijleveld and Mooijaart(2003).

$\{Y_a\}$ represents a sequence of conviction patterns for discrete time periods $a=1\ldots A$. Assume there are K latent classes , with latent class membership defined by the random process C_a with Y_a depending only on $\{C_a\}$.

There are also transition probabilities $\pi^{a}{}_{dc} = p(C_a = d \mid C_{a-1} = c)$ and starting probabilities $\pi^{1}{}_{c}$ with $1 \leq c,d \leq K$

Then the joint distribution of $\{Y_a\} = P(Y_1=y_i,\ldots Y_a=y_a,\ldots Y_A=y_A)$ is

$$\sum_{c_1}\phi_{y_1|c_1}\pi_{c_1}\ \sum_{c_2}\phi_{y_2|c_2}\pi^{2}_{c_2|c_1}\ \sum_{c_3}\phi_{y_3|c_3}\pi^{3}_{c_3|c_2}\cdots\ \sum_{c_T}\phi_{y_A|c_A}\pi^{A}_{c_A|c_{A-1}}$$

$\phi_{y_a|c_a}$ is the probability of $Y_a=y_a$ given $C_a=c_a$

Multivariate latent markov model

In the previous model ,{Y_a } represents a sequence of conviction patterns for discrete time periods *a=1…A.* How many Ys do we need at each age strip? In any five year period, we would either need to simplify conviction patterns in a fixed time period, or have a very large number of possible values of Y, and therefore a very large estimation problem.

eg 10 offence groups gives 2^{10}-1 possible Ys.
We therefore replace the univariate model with a multivariate model.

We now represent the offending history of an individual in time period *t* by a series of binary indicator variables $O_a=(O_{a1},.....O_{aJ})$ where there are *J* offence groups.
As before, $\phi_{O_a|c_a}$ is the probability of $O_a=o_a$ given $C_a=c_a$. We assume local independence.

$$\phi_{O|c} = \prod_j \left(\lambda_{j|c}\right)^{O_j} \left(1-\lambda_{j|c}\right)^{1-O_j}$$ for any *age strip a*.

$\lambda_{j|c}$ is the probability that a member of latent class c is convicted of offence j.

Parameters and modelling strategy.

This model is more complex than the earlier latent class model, as we are now modelling the transitions. This model is identical to latent transition analysis (Collins & Wulgalter, 1992 MultBehRes),

For example, with 38 offence groups,9 latent classes and six time points, we have

Latent class model: 38*9 +8 = 350 parameters
Latent Transition Analysis model: 38*9+8+5*8*8 = 670 parameters

What does latent transition analysis provide?

Class profiles of offence classes
(probability of conviction in five year period for offence I given membership of class j)

Individual probabilities of class membership
(the probabilities that an individual I belongs to class k)

The transition probabilities between 10-15 and 16-20, and between 16-20 and 21-15. Different transition matrices are estimated for each of these.

The class sizes of the latent classes at each time point

Some initial LTA models

7,626 female offenders in analysis.

We use ten of the 38 offence categories as the earlier latent class analysis suggested these were most informative for female offending.

A characteristic of all latent class models is that there are multiple “local” maxima of the likelihood. This means that we need to be careful to hit the correct solution.

LTA fitted repeatedly (100 times) with random start values. Five latent class model examined.

Best solution LTA five latent classes

Class name	Theft., receiving and fraud	Theft and shoplifting	Versatile	Non-offending	Shoplifting
Violence	0.11		0.21		
Theft	0.22	0.15	0.45		
Petty theft					
Theft from meters					
Shoplifting		0.29	0.67		1.00
Fraud and forgery	0.28		0.43		
Receiving and handling	0.14		0.33		
Criminal damage			0.20		
Absconding/bail/breach			0.13		
Drugs possession			0.33		

Probability of getting one or more convictions in a five year period given class

Estimating diversity for the offending latent classes

We assign each period of offending for each offender to the latent class with the highest probability.

We then calculate the average diversity for each latent class.

	Average diversity
Theft/receiving and fraud	0.185
Theft and shoplifting	0.121
Versatile/ frequent	0.664
Shoplifting	0.085

The female class size proportions by age for the four conviction latent classes

	Age group			Diversity
	10-15	16-20	21-25	
heft/receiving and fraud	0.003	0.293	0.511	0.185
heft and shoplifting	0.716	0.391	0.096	0.121
ersatile/ frequent	0.027	0.082	0.119	0.664
hoplifting	0.254	0.234	0.273	0.085

roportion of all offenders in ample	0.335	0.719	0.436

Female conviction transitions for those ever convicted Age group 1 (10-15) to age group 2 (16-20)

Age 16-20

		Theft/receiving and fraud	Theft and shoplifting	Versatile/ frequent	Non-offending	shoplifting
ge 0-15	Theft/receiving and fraud	1.00	0.00	0.00	0.00	0.00
	Theft and shoplifting	0.03	0.83	0.12	0.02	0.00
	Versatile/ frequent	0.04	0.22	0.56	0.18	0.01
	Non-offending	0.30	0.00	0.04	0.41	0.25
	Shoplifting	0.00	0.94	0.02	0.01	0.02

Calculated from offending sample.

Some stability observed in the four offending groups. Changes in proportions come from non-offenders joining the shoplifting and theft/receiving groups. Very little desistance observed.

Female conviction transitions for those ever convicted
Age group 2 to age group 3

	Age 21-25					
		Theft/receiving and fraud	Theft and shoplifting	Versatile/ frequent	Non-offending	shoplifting
ge 6-20	Theft/receiving and fraud	0.14	0.10	0.00	0.73	0.02
	Theft and shoplifting	0.03	0.07	0.01	0.87	0.01
	Versatile/ frequent	0.08	0.01	0.58	0.29	0.05
	Non-offending	0.61	0.00	0.05	0.00	0.34
	Shoplifting	0.04	0.00	0.02	0.87	0.07

Calculated from offending sample.

Again, either stability or desistance are the most likely outcomes for the four offending classes.

Adjusting the transitions to allow for non-offenders.

The transition matrices were estimated from those offending between ages 10-25. To estimate the transition matrix for the whole female population, we need to add in the never-offenders to row 4.

The female population of England and Wales convicted of an offence between 10-25 is estimated from our sample to be 99,051.

The total female population of England and Wales in the five birth cohorts when aged 10 is estimated to be 1,800,728. Thus there are just over 1,700,000 cases excluded.

This adds a large number of cases to the non-offending→non-offending transition.

Female conviction transitions for all females. Age group 2 to age group 3

Age 21-25

		Theft/receiving and fraud	Theft and shoplifting	Versatile/ frequent	Non-offending	shoplifting
ge 6-20	Theft/receiving and fraud	0.14	0.10	0.00	0.73	0.02
	Theft and shoplifting	0.03	0.07	0.01	0.87	0.01
	Versatile/ frequent	0.08	0.01	0.58	0.29	0.05
	Non-offending	0.01	0.00	0.00	0.98	0.01
	Shoplifting	0.04	0.00	0.02	0.87	0.07

Commentary

Other work by criminologists gives colour to latent trajectory concepts.
For example, Moffitt(1993) has suggested three groups of offenders:

Adolescent limited

Shoplifters at 16-20 will most likely stop (87% chance) but have a one in twelve chance of continuing.

Theft/receiving and fraud, and theft and shoplifting groups are similar (73% and 87% chance of stopping) – with low chances of transiting into other offending classes.

Chronic

Versatile/frequent will most likely continue in their own group (58% chance) but have a 29% chance of stopping.

Late starters

Late bloomers (Bushway, 2008) will tend to join the theft/fraud latent class and a lower chance of becoming versatile.

Commentary on specialisation.

1. The lifestyle groups found have varying degrees of diversity. Some involve themselves in only one offence, others in two or three offences, and yet others in a larger number.

2. About 10% of the female offending sample at age 16 can be considered to be truly diverse.

3. There is little evidence of offence switching in this female sample.

4. Offenders either stay in their same lifestyle class or desist.

Thus, the picture on specialisation is that we observe both specialisation and diversity in what female offenders do, and observe stability in criminal lifestyle over time.

Conclusions

- Need for a new conceptualisation of specialisation – looking at successive offences (the Forward Specialisation Coefficient approach) does not answer the important research questions.

- Lifestyle specialisation provides a more nuanced view of criminal activity over time, and can embrace both diversity and stability.

- However, lifestyle typologies will need to be determined through replication from various types of data (self-report, administrative)

- Male data will be even more challenging – more cases, more variety in offending patterns.

To conclude:

  - LTA gives real insight into the thorny problem of short and long-term specialisation in criminal behaviour.

References

Bartolucci,F. Pennoni, F and Francis, B. (2007) J Royal Statist Soc. Series A (offending transitions)

Francis, B, Liu, J and Soothill, K. (2010) Female specialisation in England and Wales. International Criminal Justice Review